

ABSTRACT OF THE DISCLOSURE

A distributed data clustering system having an integrator and at least two computing units. Each computing unit is loaded with common global parameter values and a particular local data set. Each computing unit then generates local
5 sufficient statistics based on the local data set and global parameter values. The integrator employs the local sufficient statistics of all the computing units to update the global parameter values.